

IMAGE DENOISING: POINTWISE ADAPTIVE APPROACH

POLZEHL, J. AND SPOKOINY, V.G.

*Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin
and*

*Humboldt University, Institute of Mathematics,
Unter den Linden 6, 10099 Berlin*

ABSTRACT. The paper is concerned with the problem of image denoising for the case of grey-scale images. Such images consist of a finite number of regions with smooth boundaries and the image value is assumed piecewise constant within each region. New method of image denoising is proposed which is adaptive (assumption free) to the number of regions and smoothness properties of edges. The method is based on a pointwise image recovering and it relies on an adaptive choice of a smoothing window. It is shown that the attainable quality of estimation depends on the distance from the point of estimation to the closest boundary and on smoothness properties and orientation of this boundary. It is also shown that the proposed method provides a near optimal rate of the edge estimation.

1. Introduction

One of the main problem of image analysis is the reconstruction of an image (a picture) from noisy data. It has been intensively studied last years, see e.g. the books of Pratt (1978), Grenander (1976, 1981), Rosenfeld and Kak (1982), Blake and Zisserman (1987), Korostelev and Tsybakov (1993). There are two special features related to this problem. First, the data is two-dimensional (or multidimensional). Second, the image is usually composed of several regions with rather sharp edges. Within each region the image preserves a certain degree of uniformity while on the boundaries between the regions it has considerable changes. This leads to the edge estimation problem.

1991 *Mathematics Subject Classification.* 62G07; Secondary 62G20.

Key words and phrases. edge, image, pointwise adaptive estimation, rate of estimation, regression, averaging window, testing window .

The research on this paper was carried out within the Sonderforschungsbereich 373 at Humboldt University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

The thoughtful comments of two anonymous referees, leading to significant improvement of this paper, are gratefully acknowledged. The authors thank A.Juditski, O. Lepski, A.Tsybakov and Yu.Golubev for important remarks and discussion.

A large variety of methods has been proposed for solving the image and edge estimation problem in different contexts. The most popular methods of image estimation are based on the Bayesian approach for certain parametric image modeling technique, see Haralick (1980), Geman and Geman (1984), Ripley (1988) among other. Some nonparametric methods based on penalization and regularization have been developed in Titterton (1985), Shiau, Wahba and Johnson (1986), Mumford and Shah (1989), Girard (1990).

The edge detection methods mostly do not assume any underlying parametric model. Methods based on kernel smoothing with a special choice of kernels have been discussed in Pratt (1978), Marr (1982), Lee (1983), Huang and Tseng (1988), Müller and Song (1994).

Tsybakov (1989) evaluated the optimal rate of nonparametric estimation of an image when it possesses the structure of a boundary fragment. The method allows a direct image estimation and it applies both for regular (equidistant) and random design. It is also computationally straightforward. The only inconvenience for practical applications is that it assumes an image of a special structure and some information about the image contrast is required. This method leads also to a suboptimal rate of edge detection $(n^{-1} \log n)^{1/2}$.

A general asymptotic minimax theory of edge estimation has been developed in Korostelev and Tsybakov (1993). For instance, they showed that linear methods are not optimal for images with sharp edges. Imposing some smoothness restrictions on the boundary, they found the minimax rate $n^{-\gamma/(\gamma+1)}$ of edge estimation, γ being the degree of edge smoothness, and constructed rate-optimal estimators for images with the structure of a boundary fragment. The proposed methods are essentially nonlinear and they involve column-wise change-point analysis.

In the present paper, we propose another approach which is based on direct image estimation at each design point. We apply a simple linear estimator which is the average of observations over a window selected in a data-driven way. Then we study which accuracy of edge estimation is provided by this procedure. In spite of the fact that linear methods are only suboptimal in edge estimation, the results of this paper show that a non-linearity which is incorporated in the linear method by an adaptive choice of an averaging window allows to get a near optimal quality of edge recovering.

The presented approach can be viewed as one more application of the idea of pointwise adaptive estimation, see Lepski (1990, 1992), Lepski, Mammen and Spokoiny (1997), Lepski and Spokoiny (1997), Spokoiny (1996). In the last paper a pointwise adaptive procedure was applied to estimate a function with heterogeneous smoothness properties, allowing, for instance, jumps or jumps of derivatives. The methods based on pointwise (or local) adaptation are especially fruitful in situations where a complex object like a function with heterogeneous smoothness properties admits a simple description in a small neighborhood of each point and the procedure, being applied at this point, adapts exactly to the underlying local structure. In essence, the procedure searches for a largest local vicinity of the point of estimation where the local structural assumption fits well to the data.

We now apply this idea to the problem of image estimation. We focus on the case of piecewise constant images i.e. we assume that the image consists of a finite number of regions and the image value is constant within each region. The image is

observed with noise on a regular grid in the unit square and we estimate the image value separately at each design point via a data-driven choice of the averaging window. The benefit of this approach is that it is very general in nature and it does not require to specify the number of regions, difference between values of the image function f for different regions or regularity of each edge. Moreover, this method can be applied to estimate any function which can be well approximated by a constant function in a local vicinity of each point.

We consider the regression model

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n, \quad (1.1)$$

where $X_i \in [0, 1]^d$, $i = 1, \dots, n$, are given design points and ξ_i are individual independent random errors. Below we will suppose that ξ_i , $i = 1, \dots, n$, are i.i.d. $\mathcal{N}(0, \sigma^2)$ with a given noise level σ .

Next we suppose that the cube $[0, 1]^d$ is split into M regions A_m , $m = 1, \dots, M$ each of them is a connected set with an edge (boundary) G_m . The function f is assumed constant within each region A_m , i.e.

$$f(x) = \sum_{m=1}^M a_m \mathbf{1}(x \in A_m) \quad (1.2)$$

where a_1, \dots, a_M are unknown constants. The problem is to estimate the image function $f(x)$ or, equivalently, to estimate the values a_1, \dots, a_M and to decide for each point X_i what the corresponding region is.

The idea of the proposed method is quite simple. We search for a maximal possible window U containing x^0 in which the function f is well approximated by a constant. Further this constant is taken as the resulting estimate. Of course, the choice of the considered class of windows plays the key role for such an approach. We will discuss this problem a little bit later. Now we suppose for a moment that we are given a class \mathcal{U} of windows U , each of them being a subset of the unit cube $[0, 1]^d$ containing the point of interest x^0 . By N_U we denote the number of design points in U . The assumption that f is constant in U leads to the obvious estimator \hat{f}_U of $f(x^0)$ which is the mean of observations Y_i over U .

To characterize the quality of the window U we calculate the residuals $\varepsilon_{U,i} = Y_i - \hat{f}_U$ and we test the hypothesis that these residuals $\varepsilon_{U,i}$ can be treated within the window U as a pure noise. Finally the procedure selects the maximal (in number of points N_U) window for which this hypothesis is not rejected.

The paper is organized as follows. In the next section we present the procedure, Section 3 contains the results describing the quality of this procedure. In Section 4 we specify the general results to the case of an equidistant design and discuss the problem of edge estimation in this context. Section 5 contains some simulated examples. The proofs are mostly deferred to Section 6.

2. Estimation Procedure

Let data Y_i, X_i , $i = 1, \dots, n$ obey model (1.1). We will estimate $f(x^0)$ for a given x^0 . Typically x^0 is a design point i.e. the image is recovered at the same points

where it is observed. Note, nevertheless, that the procedure applies in general for any point x^0 and can be used for image interpolation as well.

Given a family of windows \mathcal{U} and $U \in \mathcal{U}$, set N_U for the number of the points X_i falling in U ,

$$N_U = \#\{X_i \in U\}.$$

We will suppose that $N_U \geq 2$ for each $U \in \mathcal{U}$. We assign to each $U \in \mathcal{U}$ the estimator \hat{f}_U of $f(x^0)$ by

$$\hat{f}_U(x^0) = \hat{f}_U = \frac{1}{N_U} \sum_U Y_i.$$

Here the sum over U means the sum over design points in U .

Our adaptation method is based on the analysis of the residuals $\varepsilon_{U,i} = Y_i - \hat{f}_U$. We introduce another family $\mathcal{V}(U)$ of windows V , each of them is a subwindow of U , i.e. $V \subset U$. We require only that $N_V := \#\{X_i \in V\} \geq 2$ for all $V \in \mathcal{V}(U)$. One example of the choice of the families \mathcal{U} and $\mathcal{V}(U)$ for the case of the equidistant design is presented in Section 4.

By C_U we denote the cardinality of $\mathcal{V}(U)$,

$$C_U = \#\mathcal{V}(U), \quad (2.1)$$

and by C^* the value

$$C^* = \max \left\{ \#\mathcal{U}, \max_{U \in \mathcal{U}} C_U \right\}.$$

For each $V \in \mathcal{V}(U)$ set

$$T_{U,V} = \frac{1}{\sigma_{U,V} N_V} \sum_V \varepsilon_{U,i} = \frac{1}{\sigma_{U,V} N_V} \sum_V (Y_i - \hat{f}_U) = \frac{\hat{f}_V - \hat{f}_U}{\sigma_{U,V}}$$

where \sum_V means summation over the index set $\{i : X_i \in V\}$ and $\sigma_{U,V}$ is the standard deviation of the difference $\hat{f}_V - \hat{f}_U$,

$$\sigma_{U,V}^2 = \sigma^2 \frac{N_U - N_V}{N_U N_V} = \sigma^2 (N_V^{-1} - N_U^{-1}).$$

Define now

$$\varrho_{U,V} = \mathbf{1}(|T_{U,V}| > t^*)$$

where

$$t^* = \sqrt{2\lambda + 2 \log C^*} \quad (2.2)$$

and λ is some constant which determines the probability of wrong classification.

We say that U is rejected if $\varrho_{U,V} = 1$ for at least one $V \in \mathcal{V}(U)$ i.e. if $\varrho_U = 1$ with

$$\varrho_U = \sup_{V \in \mathcal{V}(U)} \varrho_{U,V} = \mathbf{1} \left(\sup_{V \in \mathcal{V}(U)} |T_{U,V}| > t^* \right).$$

The adaptive procedure selects among all non-rejected U from \mathcal{U} one which maximizes N_U ,

$$\hat{U} = \operatorname{argmax}_{U \in \mathcal{U}} \{N_U : \varrho_{U,V} = 0 \text{ for all } V \in \mathcal{V}(U)\}.$$

If there is more than one non-rejected set U attaining the maximum, then any of them can be taken. Finally we set

$$\hat{f}(x^0) = \hat{f}_{\hat{U}}(x^0) = \hat{f}_{\hat{U}}.$$

3. The main results

Below we describe some properties of the proposed estimation procedure and state the result about the corresponding accuracy of estimation.

Let x^0 be a given point. Our target is the image value $f(x^0)$. In the sequel we assume that x^0 is from region A_m for some $m \leq M$. Of course, the number m , as well as the total numbers of regions M and the structure of each region are unknown.

We also suppose we have a family \mathcal{U} of windows containing x^0 and for each $U \in \mathcal{U}$ a family of testing subwindows $\mathcal{V}(U)$.

Our results are stated using the notion of an “ideal” window U^* from \mathcal{U} . Namely, let U^* stand for the largest (in the number of design points) window from \mathcal{U} which is contained in A_m . It is also assumed that this set is not empty. Otherwise we set simply $U^* = \emptyset$ and $N_{U^*} = 0$.

To state the results about the quality of estimation by our adaptive procedure, we need some conditions on the families \mathcal{U} and $\mathcal{V}(U)$, $U \in \mathcal{U}$.

(U) :

- (U.1) every set U from \mathcal{U} contains x^0 ;
- (U.2) there is an integer number K such that for every $U, U' \in \mathcal{U}$, the intersection $U \cap U'$ contains a testing window $V \in \mathcal{V}(U)$ with

$$N_V \geq K;$$

- (U.3) Let U^* be the maximal window from \mathcal{U} which is contained in A_m . Let then $U \in \mathcal{U}$ be such that $N_U > N_{U^*}$. If the difference $U \setminus A_m$ does not contain any $V \in \mathcal{V}(U)$ with $N_V \geq K$, then there is a $V_1 \in \mathcal{V}(U)$ such that $V_1 \subseteq A_m$ and

$$N_{V_1} \geq N_{U^*}/2.$$

The choice of the constant K in (U.2) will be discussed later.

Remark 3.1. Condition (U.2) from the above did not appear in the univariate case, see Spokoiny (1996), but it is rather essential in the multivariate situation. This can be illustrated by the following example. Let the point of interest x^0 lie in A_1 but near the boundary with another region, say A_2 . Let also there be two windows U_1 and U_2 from \mathcal{U} such that all design points from U_1 are in A_1 and almost all design points from A_2 except x^0 are in A_2 . Both these window with a large probability will not be rejected and if, in addition, $N_{U_2} > N_{U_1}$, then the procedure selects U_2 rather than U_1 . This will lead to a inconsistent estimator $\hat{f}(x^0) \approx a_2$

obtained by averaging over U_2 . We shall see that condition (U.2) does not allow such a situation.

Conditions (U.1) and (U.2) rely only on the structure of the set \mathcal{U} of considered windows whereas condition (U.3) rely also on the properties of the edge of the region A_m containing x^0 . This condition can be commented in the following way. Let U^* be an “ideal” window and let U be another window such that $N_U > N_{U^*}$. This means that U is “large” compared with U^* . If the difference $U \setminus A_m$ is small in the sense that it does not contain any testing window V with $N_V \geq K$, then the intersection $U \cap A_m$ is large in the sense that it contains a window V with $N_V \geq N_{U^*}/2$.

Now we are in a position to state the main result.

Theorem 3.1. *Let the image function $f(x)$ be piecewise constant, see (1.2). Let some $K > 0$ be fixed and let conditions (U.1) through (U.3) be satisfied. Let also*

$$C^* = \max \left\{ \#\mathcal{U}, \max_{U \in \mathcal{U}} \#\mathcal{V}(U) \right\}$$

and $t^* = \sqrt{2\lambda + 2 \log C^*}$. If $x^0 \in A_m$ and if for all $m' \neq m$, it holds

$$|a_m - a_{m'}| \geq 6\sigma t^* K^{-1/2}, \quad (3.1)$$

then

$$P_f \left(\left| \hat{f}(x^0) - f(x^0) \right| > 2\sigma (N_{U^*}/2)^{-1/2} t^* \right) \leq 3e^{-\lambda}.$$

3.1. The accuracy of estimation

The result of Theorem 3.1 guarantees a reasonable quality of estimation only in the case when the “ideal” window U^* is sufficiently large. In particular, if there is no window in \mathcal{U} contained in the same region A_m as x^0 , then this result does not apply. From the other side, if x^0 is an internal point of the region A_m and if the region is comparable in size with the whole square $[0, 1]$, then typically there are large windows with of order n points inside, that is, $N_{U^*} \asymp n$. Therefore, inside each “large” region, the proposed procedure estimates the image value with the rate $n^{-1/2}$ up to some logarithmic term. Finally, if x^0 lies near the boundary of the region A_m , then the size of U^* depends on the distance of x^0 to the boundary of A_m and on the smoothness properties of this boundary. The same is valid for the quality of estimation. More detailed discussion can be found in Section 4.3.

The notion of an “ideal” window $U^* = U^*(x)$ can be regarded as a multi-dimensional analog of the notion of a local smoothness characteristic, cf. Lepski, Mammen and Spokoiny (1997), Lepski and Spokoiny (1997) or Spokoiny (1996). Donoho and Johnstone (1996) used the notion of an “oracle” to explain a similar thing. Indeed, if we knew from an oracle the “ideal” window U^* , then we define our estimate by averaging over U^* . This estimate can also be called “ideal” or “oracle” estimate. Our main results claim that our procedure is locally adaptive in the sense that it delivers the same rate of accuracy $N_{U^*}^{-1/2}$ (up to a log-term) as if the “ideal” windows were known.

Now we shortly discuss the choice of the parameter K entering in condition (U.2) and (3.1). First we mention that the value C^* is typically of order n^r with

some positive r . Therefore, the value $t^* = \sqrt{2\lambda + 2\log C^*}$ is of order $\sqrt{\log n}$. If the image contrasts $a_m - a_{m'}$ are all of order 1, then it follows for K from (3.1)

$$K \geq \text{constant } \sigma^2 \log n.$$

Therefore, K has to be of logarithmic order but its choice depends on the noise level and on the image contrast. For more discussion see also Section 4.3.1 and Section 5.

4. The case of an equidistant design

In this section we specify our procedure and results to the case of a regular equidistant design in the unit square $[0, 1]^2$. We also discuss the problem of edge estimation.

Suppose therefore that we are given n design points X_1, \dots, X_n with $X_i = (X_{i,1}, X_{i,2}) \in [0, 1]^2$. Without loss of generality we may assume that \sqrt{n} is an integer and denote $\delta = n^{-1/2}$. Now each design (or grid) point X_i can be represented in the form $X_i = (k_1\delta, k_2\delta)$ with nonnegative integers k_1, k_2 .

As above, we consider the problem of estimating the image value at a point x^0 by observations Y_1, \dots, Y_n described by the model equation (1.1). In this section we restrict ourselves to estimation on the grid, i.e. we suppose additionally that x^0 is a grid point.

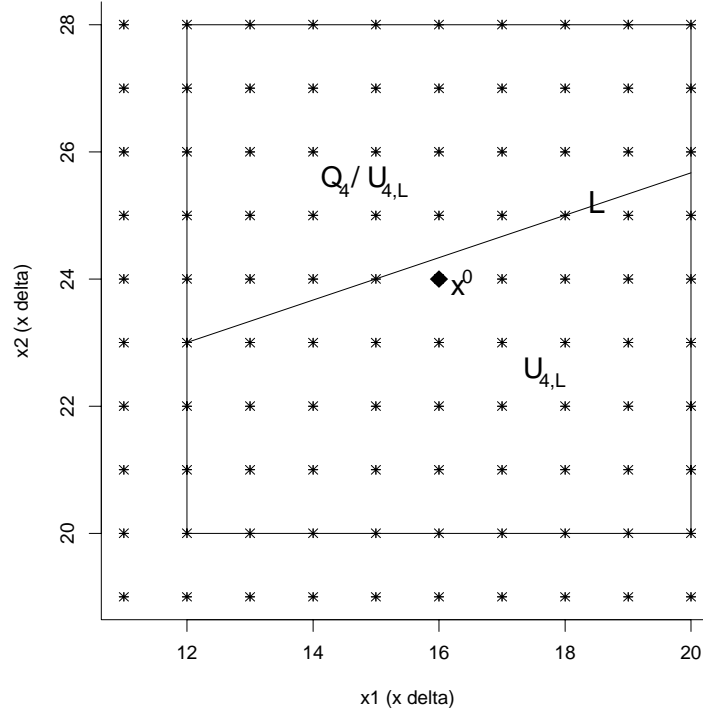
We begin by describing one possible choice of the set of windows \mathcal{U} . Then we specify the result of Theorem 3.1 to this case and consider the problem of edge estimation. Then we compare our results with the existing results from the literature.

4.1. An example of the set of windows

Our procedure involves two external parameters K and D . The integer K enters in condition (U.2) and in Theorem 3.1 and we have to ensure in our construction that the number of design points in the intersection of every two windows of the constructed family is at least K . The parameter D controls the maximal size of the considered windows.

Let, given an integer $d > 1$, Q_d be the axis-parallel square with the centre at x^0 and with side length $2d\delta$, δ being $n^{-1/2}$. Obviously Q_d contains exactly $(2d+1)^2$ design points. First we describe all windows associated with this square. Denote for every two different design points X_i, X_j by $L(X_i, X_j)$ the straight line passing through these points. If X_i and X_j belong to Q_d , then this line splits the square into two parts. We define $U_{d,L}$ as one of them which contains x^0 . For a formal description, we represent the line L by the equation $a_L^T x + b_L = 0$ where a_L is a vector in \mathbb{R}^2 and b_L is a real number. Then the sign of the expression $a_L^T x + b_L$ determines the position of a point x w.r.t. the line L . In particular, all the points from the same side of L have the same sign for this quantity. Now we set

$$U_{d,L} = \{x \in Q_d : \text{sign}(a_L^T x + b_L) \text{sign}(a_L^T x^0 + b_L) \geq 0\},$$

FIGURE 1. Q_4 and one selection of U_L for $x = (16\delta, 23\delta)$.

see Figure 1.

This definitions means that the points lying on the line L are also included in $U_{d,L}$. For the sequel, it is convenient to identify two windows $U_{d,L}$ and $U_{d,L'}$ if they contain the same subset of design points from Q_d . We then define a set \mathcal{U}_d of windows associated with the square Q_d by

$$\mathcal{U}_d = \{U_{d,L}, \forall L\}.$$

In other words, we obtain the family \mathcal{U}_d considering splits of the square Q_d into two parts and taking each time the largest from these two. Clearly the equivalent definition is

$$\mathcal{U}_d = \{U_{d,L}, L = L(X_i, X_j), X_i, X_j \in Q_d\}.$$

We also denote by $\tilde{\mathcal{U}}_d$ the set of all splits including the smallest parts too,

$$\tilde{\mathcal{U}}_d = \{U_{d,L}, Q_d \setminus U_{d,L}, L = L(X_i, X_j), X_i, X_j \in Q_d\}.$$

The total number of windows in $\tilde{\mathcal{U}}_d$ can be very roughly estimated by $(2d+1)^4/2$.

Very roughly we define our set \mathcal{U} as the union of \mathcal{U}_d with different d . But we have also to provide condition (U.2), and for this it is required to restrict slightly this class. Given a line L , denote by L' the parallel line passing through x^0 , so that $L' = \{x : a_L^T(x - x^0) = 0\}$. By $K_{d,L}$ we denote the number of design points in Q_d between L and L' ,

$$K_{d,L} = \{X_i \in U_{d,L} : \text{sign}(a_L^T(X_i - x^0)) \text{sign}(a_L^T(x_L - x^0)) \geq 0\}$$

where x_L denotes an arbitrary point on L . Note that the points lying on both lines L and L' are included. This definition is illustrated in Figure 2. Finally we set

$$\mathcal{U}_{d,K} = \{U_{d,L} \in \mathcal{U}_d : K_{d,L} \geq K\}.$$

Lemma 4.1. *The intersection of every two windows from $\mathcal{U}_{d,K}$ contains at least K design points.*

Proof. Let U_{d,L_1} and U_{d,L_2} be two windows from $\mathcal{U}_{d,K}$ corresponding to some lines L_1 and L_2 respectively. Let also L'_1 be the parallel to L_1 line passing through x^0 .

For each grid point X_i we denote by X_i^s the symmetric to X_i point w.r.t. x^0 , see Figure 2. Now we may use the simple fact that if X_i lies between L_1 and L'_1 , then X_i^s belongs to U_{d,L_1} , and that either X_i or X_i^s belongs to U_{d,L_2} . \square

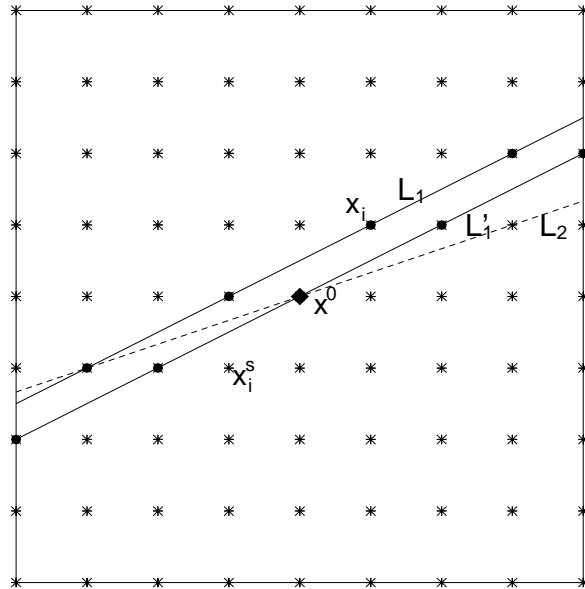


FIGURE 2. L , L' and L_2 .

For every window U from $\mathcal{U}_{d,K}$ we define a family of testing windows $\mathcal{V}_d(U)$ in a similar manner. Namely we take all windows V from $\tilde{\mathcal{U}}_d$ which are contained in U , and also all differences $U \setminus V$,

$$\mathcal{V}_d(U) = \{V \text{ or } U \setminus V : V \in \tilde{\mathcal{U}}_d, V \subset U\}.$$

Next we introduce a set \mathcal{D} of all considered d -values. This set is a subset of the index set $\{1, 2, \dots, D\}$ and hence it contains at most D elements. Now we define

$$\mathcal{U} = \mathcal{U}_{\mathcal{D},K} = \bigcup_{d \in \mathcal{D}} \mathcal{U}_{d,K}$$

For every U from \mathcal{U}_d and each $d' < d$, let $U_{d'} = U \cap Q_{d'}$. Then we set

$$\mathcal{V}(U) = \bigcup_{d' \in \mathcal{D}: d' \leq d} \mathcal{V}_{d'}(U_{d'}).$$

For the above defined set \mathcal{U} , condition (U.1) is fulfilled by construction. Now we are checking (U.2).

Lemma 4.2. *The sets \mathcal{U} and $\mathcal{V}(U)$, $U \in \mathcal{U}$, fulfill (U.2).*

Proof. Let $U = U_{d,L}$ and $U' = U_{d',L'}$ be two windows from \mathcal{U} . Without loss of generality we may assume that $d = d'$. Otherwise, if for instance $d' > d$, we take $U' \cap Q_d$ instead of U' . We also restrict ourselves to the most interesting case when the lines L and L' do not intersect within Q_d . Then by construction $V = U \cap U'$ is a testing window from $\mathcal{V}(U)$ and by Lemma 4.1 it contains at least K design points. \square

It is obvious that the total number of windows in $\mathcal{U} = \mathcal{U}_{\mathcal{D},K}$ is bounded by

$$\sum_{d=1}^D (2d+1)^4/2 \leq (2D+1)^5/10$$

and the same is valid for the cardinality of every $\mathcal{V}(U)$ with $U \in \mathcal{U}$. Therefore $C^* \leq (2D+1)^5/10$. Hence the value t^* from (2.2) can be roughly estimated as follows:

$$t^* = \sqrt{2\lambda + 2 \log C^*} \leq \sqrt{2\lambda + 10 \log(2D+1)}.$$

Further we discuss the properties of the estimate $\hat{f}(x^0)$ corresponding to the previously described sets \mathcal{U} and $\mathcal{V}(U)$, $U \in \mathcal{U}$.

We begin with the very simple situation when the point of interest x^0 lies inside a homogeneous region A_m . We will see that such a situation the value $f(x^0)$ is estimated with the rate $n^{-1/2}$ up to a logarithmic term.

Theorem 4.1. *Let the point x^0 belong to a region A_m together with the square*

$$Q_\epsilon(x^0) = \{x : \max\{|x_1 - x_1^0|, |x_2 - x_2^0|\} \leq \epsilon\}.$$

Let also K satisfy the condition

$$K \geq (4t^* \sigma \alpha^{-1})^2 \tag{4.1}$$

with $\alpha = \max\{|a_m - a_{m'}|, m' \neq m\}$ and

$$t^* = \sqrt{2\lambda + 10 \log(2n^{1/2}\epsilon)}. \tag{4.2}$$

If $D \geq Cn^{1/2}\epsilon$ with some positive constant $C \leq 1$, then

$$\mathbf{P}_f \left(|\hat{f}(x^0) - f(x^0)| \geq 2\sigma t^* (2Cn^{1/2}\epsilon)^{-1} \right) \leq 3e^{-\lambda}.$$

This result is a corollary of Theorem 3.1. It suffices to note that the window U coinciding with the square Q_D belongs to the family \mathcal{U} and it is contained in A_m . Hence $N_{U^*} \geq N_U \geq 4D^2$. Note also that here x^0 should not necessary be a grid point.

Remark 4.1. It is worth to mention that in this situation much simpler methods apply as well, see e.g. Tsybakov (1989).

4.2. The accuracy of estimation near an edge

Now we are going to apply Theorem 3.1 to the case when the point of interest x^0 lies near an edge of the corresponding region. First we illustrate the importance of a careful estimation near an edge by the following

Example 4.1. Let A be a circle inside the unit square with a radius $r > 0$. We do not suppose that the center of this circle is at a grid point. The radius r may be also arbitrary. We set $\rho = C/n$ with some constant $C > 1$ and consider a band of width ρ near the edge of A . Note that this width is essentially smaller than the grid step $\delta = n^{-1/2}$, if C is not too large. The Lebesgue measure of this band is about $2\pi r \rho$, so, for the uniform random design, the number of design points inside this band would be in mean about $2\pi r \rho n = 2\pi r C$. It can be shown by using the arguments from the theory of continuous fraction, see Khintchine (1949) or Lemma 5.2 below, that under the equidistant design, we have essentially the same (in order) number of design points inside this band. On the other side, it is well known, that the quality of estimation near an edge is especially important by visualization. This is illustrated in Figure 3. Even single errors in image segmentation are visible and they lead to a significant deterioration of the image, see e.g. Figure 4.

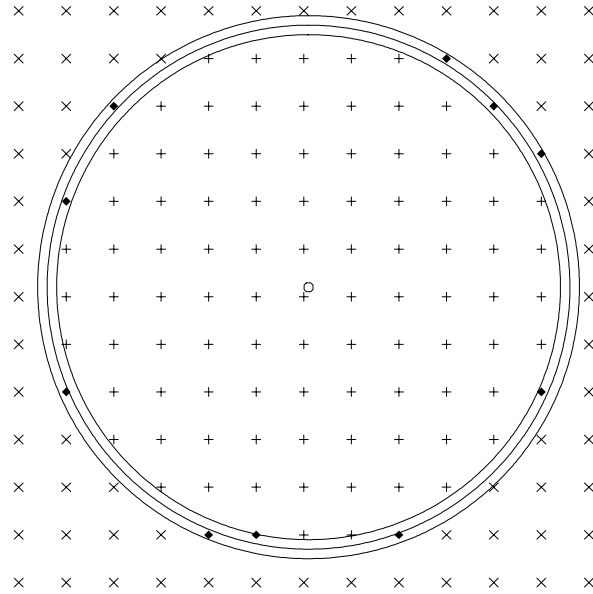


FIGURE 3. Band of width $\rho = .4\delta$ around a dislocated circle of radius 5.5δ . The band contains 10 points ($2\pi r \rho n \approx 13.8$).

Let x^0 belongs to a region A_m and lies near its the edge G with another region $A_{m'}$. We assume also that this edge is regular in the sense that it can be well approximated by a straight line in some small vicinity of the point x^0 .

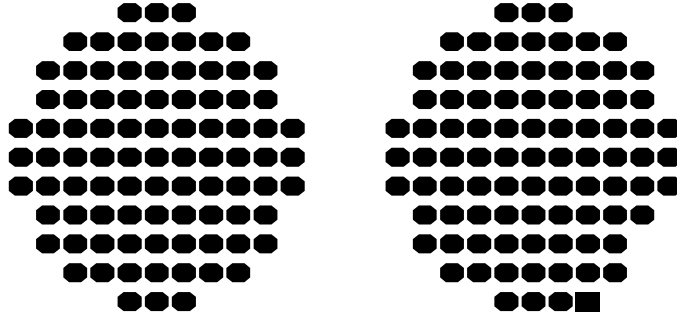


FIGURE 4. Ball of radius 5.5δ on a grid (left) and same ball with 2 misspecified pixels near the edge.

Without loss of generality we may assume that the edge G can be parametrized in a neighborhood of the point x^0 by the equation $x_2 = g(x_1)$ with some differentiable function g and that $|g'(x_1^0)| \leq 1$. (Otherwise another parametrization of the form $x_1 = g(x_2)$ is to be used.) Now the image function f at least in a neighborhood of the point x^0 can be represented in the form

$$f(x) = \begin{cases} a_{m'} , & x \in A_{m'} = \{x_2 > g(x_1)\}, \\ a_m , & x \in A_m = \{x_2 \leq g(x_1)\}. \end{cases} \quad (4.3)$$

The distance from x^0 to the edge G of A_m can be characterized by the value $g(x_1^0) - x_2^0$, see Figure 5.

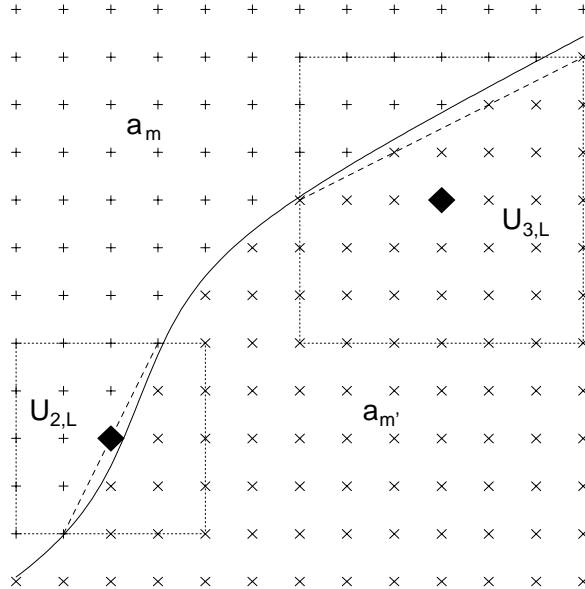


FIGURE 5. Optimal sets Q_d and $U_{d,L}$ for 2 points x^0 .

In the next result we suppose that the edge function g is smooth in the sense that it belongs to the Hölder class $\Sigma(\gamma, P)$ with some parameters $\gamma \in (1, 2]$ and $P > 0$. This means that g fulfills the condition

$$|g'(s) - g'(t)| \leq P|s - t|^\gamma, \quad \forall s, t.$$

We consider the properties of the estimate $\hat{f}(x^0)$ for this situation assuming that the value D is sufficiently large and $\mathcal{D} = \{1, 2, \dots, D\}$.

Theorem 4.2. *Let the image function $f(x)$ be of the form (4.3) in a neighborhood $Q_\epsilon(x^0)$ of the point x^0 with some positive $\epsilon > 0$ and let a grid point x^0 belong to A_m . The function g describing the edge G is supposed to be in the Hölder class $\Sigma(\gamma, P)$. Let K and t^* satisfy the conditions (4.1) and (4.2). If D fulfills $D \leq \epsilon n^{1/2}$ and*

$$D \geq d_\gamma := \left\lceil n^{\frac{\gamma-1}{2(\gamma+1)}} (K/P)^{\frac{1}{\gamma+1}} \right\rceil,$$

$[a]$ being the integer part of a , and if the distance $g(x_1^0) - x_2^0$ satisfies

$$g(x_1^0) - x_2^0 \geq 2P^{\frac{1}{\gamma+1}} (n/K)^{-\frac{\gamma}{\gamma+1}},$$

then

$$P_f \left(|\hat{f}(x^0) - f(x^0)| \geq 2\sigma t^* d_\gamma^{-1} \right) \leq 3e^{-\lambda}.$$

4.3. Edge estimation

Now we shortly discuss the problem of edge estimation. Note that the above described procedure is assigned for estimating the image function f and there is no edge estimation subroutine. Nevertheless, in the case of an image with the structure of a boundary fragment, the procedure estimates the value $f(x^0)$ consistently and even with some rate if the point x^0 is bounded away from the edge with the distance of order $\psi_\gamma(n) = (n^{-1} \log n)^{\frac{\gamma}{\gamma+1}}$. The minimal distance between the point x^0 and the edge G which is sufficient for consistent estimation of $f(x^0)$ can be regarded as the accuracy of edge estimation. Indeed, due to Theorem 4.2 we get a consistent image estimation outside of the “band of insensitivity” of the width of order $\psi_\gamma(n)$ and this estimator can be used for edge restoration.

Now we aim to compare this accuracy with the earlier results on edge estimation. The problem of the edge estimation was considered in details in Korostelev and Tsybakov (1993). They have shown that the rate of edge estimation critically depends on the smoothness properties of the function g defining the edge. In particular, if g belongs to a Hölder class $\Sigma(\gamma, 1)$, then the accuracy of edge estimation, being measured in the Hausdorff metric, is $(n/\log n)^{-\gamma/(\gamma+1)}$. We see that our procedure provides essentially with the same rate. Note meanwhile, that Korostelev and Tsybakov (1993) stated their results under a random or jittered design, see p.92 there. Under the regular design, the rate of edge estimation is equal to the grid step $\delta = n^{-1/2}$, Korostelev and Tsybakov (1993, p.99). This can be illustrated by the following example: if the edge G is a straight horizontal line, then for any shift of this line within an interval between two neighbor grid lines, we have the same distribution on the space of observations and hence the accuracy of

estimation cannot be better (in rate) than $n^{-1/2}$. Only assuming a random design, the above mentioned improving in the rate of edge estimation is possible.

We proceed under the regular design but we estimate the value of the image at a grid point. We see that this fact also allows us to get a better accuracy of estimation $\psi_\gamma(n) = (n/\log n)^{-\gamma/(\gamma+1)}$ with $\gamma \in (1, 2]$.

The result of Theorem 4.2 delivers some additional information about dependence of the quality of edge estimation on the noise level σ , the image contrast α and the orientation of the edge G described by the value $z = g'(x_1^0)$.

4.3.1. Accuracy versus noise level and image contrast

We see from the result that for consistent estimation, the required distance from the point of interest x^0 to the edge G should be of order $(n^{-1}K P)^{-\frac{\gamma}{\gamma+1}}$. This expression depends on the noise level σ only through K which must fulfill $K \geq C\sigma^2\alpha^{-2}\log n$ with some constant C . We see that when the noise level increases the quality of edge recognition decreases by the factor $\sigma^{2\gamma/(\gamma+1)}$. Another possible description of this influence is to say that increasing in the noise level is equivalent to increasing the grid step $n^{-1/2}$ by the factor σ^{-1} .

All this remains valid for dependence of the quality of estimation on the value of image contrast $\alpha := \max\{|a_m - a_{m'}|, m' \neq m\}$. The only difference is that this dependence is with another sign: when the contrast increases the quality increases as well, ad wise versa. Both these issues are in accordance with the one-dimensional case (Spokoiny 1996) and with similar results for a random design (Mammen and Tsybakov 1995).

4.3.2. Accuracy versus edge orientation

The previous results are completely analogous to the existing results on image recognition for a random design, cf. Korostelev and Tsybakov (1993). Now we discuss shortly a problem which appears only for the regular design. Namely it is a dependence of the quality of edge estimation on the edge orientation. This orientation is characterized by the value $z = g'(x_1^0)$. By inspecting the proof one can see that the quality of estimation depends critically on the quality of approximation of z by rational numbers with bounded denominators. It follows from the result that the worst case leads just to the above indicated rate. At the same time, if z is a rational number, $z = p/q$, with a bounded q , or if z is very close to such a rational number, then the quality of estimation can be improved. We present one result in this spirit restricting ourselves to the case of a rational z .

Theorem 4.3. *Let the image function $f(x)$ be of the form (4.3) in a neighborhood $\{x : |x - x^0| \leq \epsilon\}$ of the point x^0 with some positive $\epsilon > 0$ and let a grid point x^0 belong to A_m . The function g describing the edge G is supposed to be in the Hölder class $\Sigma(\gamma, P)$ and additionally $z := g'(x_1^0) = p/q$ with some integer $p \leq q$. Let K and t^* satisfy the conditions (4.1) and (4.2). If D fulfills*

$$Kq \leq D \leq \epsilon n^{1/2}$$

and if the distance $g(x_1^0) - x_2^0$ satisfies

$$g(x_1^0) - x_2^0 \geq P(Kqn^{-1/2})^\gamma, \quad (4.4)$$

then

$$\mathbf{P}_f \left(|\hat{f}(x^0)| \geq 2\sigma t^*(Kq)^{-1} \right) \leq 3e^{-\lambda}.$$

As a corollary of this result we conclude that an edge with a rational (e.g. with horizontal or vertical) orientation can be estimated with the rate $(n^{-1} \log n)^{\gamma/2}$ which approaches n^{-1} for $\gamma = 2$.

4.3.3. Rate optimality

The next natural question is about the optimal rate of edge estimation. Korostelev and Tsybakov (1993) show that the rate $\psi_\gamma(n) = (n/\log n)^{-\gamma/(\gamma+1)}$ cannot be improved for the case of a random design. But this lower bound does not apply for the case of estimation at design points with a regular design. The next assertion shows that the accuracy $\psi_\gamma(n)$ cannot be essentially improved uniformly over the class of all boundary fragments in the case of a regular design as well.

From Theorem 4.3 we know that some improvement in the accuracy of edge estimation is still possible for images with a special orientation. We will see that the accuracy delivered by our procedure is at least near optimal in this situation too.

Let some grid point x^0 be fixed and let an image have the structure of a smooth boundary fragment (at least locally near the point x^0) with an edge G determined by a function $g = g(x_1)$ from the Hölder ball $\Sigma(\gamma, 1)$ with $\gamma \in (1, 2]$. The function g determines the image function f_g with $f_g(x) = \mathbf{1}(x_2 \geq g(x_1))$ for $x = (x_1, x_2)$. We stand also $G = G_g$ for the corresponding edge i.e. $G = \{x : x_2 = g(x_1)\}$.

We are interested in the minimal distance between the point x^0 and the edge G which allows a consistent estimation of $f(x^0)$ for image functions f of the form f_g with g from $\Sigma(\gamma, 1)$.

Theorem 4.4. *Let K, D be integers and let $z = p/q$ be a unreducible rational number with $0 \leq p \leq q$. Let then ψ_n stand for*

$$\psi_n = \min\{n^{-1/2}q^{-1}, (qKn^{-1/2})^\gamma\}.$$

Then there exist a constant $\varkappa > 0$ depending only on γ and two functions g_0 and g_1 from $\Sigma(\gamma, 1)$ such that $g'_0(x_1^0) = z$, $g'_1(x_1^0) = z$,

$$g_0(x_1^0) \geq x_2^0 + \varkappa\psi_n, \quad g_1(x_1^0) \leq x_2^0 - \varkappa\psi_n, \quad (4.5)$$

and such that for any estimator \tilde{f}

$$\mathbf{P}_{g_0} \left(|\tilde{f}(x^0) - f_{g_0}(x^0)| > 1/2 \right) + \mathbf{P}_{g_1} \left(|\tilde{f}(x^0) - f_{g_1}(x^0)| > 1/2 \right) \geq c, \quad (4.6)$$

where c is some positive number depending only on K .

Remark 4.2. If we apply this theorem with a small q , then we get the lower bound for the result of Theorem 4.3. Maximizing ψ_n with respect to q leads to the choice $q \approx K^{-\gamma/(\gamma+1)} n^{(\gamma-1)/(2\gamma+2)}$ and to the lower bound $\psi_n \approx (K/n)^{\gamma/(\gamma+1)}$ coinciding in order with the upper bound from Theorem 4.2.

5. Simulation results

The simulation results presented in this section are based on an implementation of the proposal for \mathcal{U} and $\mathcal{V}(U)$ given in Section 4.1. In the algorithm we restrict the splits L by $a_L^T = (a_{1,L}, a_{2,L})$ with $|a_{i,L}| \leq q^*$, where the parameter q^* allows to limit the complexity of the procedure. We use $q^* = 5$. The algorithm starts with $d = 0$ although not all theoretical assumptions are fulfilled for this case. We also take $D = 12$. This means that the maximal considered windows for each point contains $25 \times 25 = 625$ design points. The theoretical recommendation (2.2) for the choice of the important parameter t^* , entering in the description of the method, turns out to be far too conservative. In our calculations $t^* \approx 3$ gives reasonable results.

Besides an illustration of the feasibility of our approach, there are three main aspects we intend to study by our simulation. These are the influence of the signal/noise ratio and the dependence of bias and variance of the estimates from orientation and curvature of the function g used in the local parametrization of the edge. To study these aspects we use two simple images both characterized by a boundary function g of constant curvature. Image 1 contains 56×56 pixel ($\delta = 1/56$) with values 1 and 0 for pixels inside and outside of a circle with radius 12δ centered in the image. Image 2 is constructed similarly containing 80 pixels ($\delta = 1/80$) and using a circle of radius 24δ .

Figure 6 displays results of the reconstruction of distorted images for three situations. In each row the left image shows the distorted original, the central image gives the estimate obtained by our algorithm and the right image shows the size of the set $\hat{U}(x^0)$ used for each pixel. Note the dependence between location of x^0 with respect to the boundary and size of $U(x^0)$ which clearly illustrates the adaptivity of our procedure.

For both images 1 and 2 we conducted a small simulation study of size $n_{sim} = 100$. The distorted images are generated by adding Gaussian errors to the images. We use four values for the standard deviation $\sigma = .125, \sigma = .25, \sigma = .5$ and $\sigma = 1$ corresponding to signal/noise ratios of 8, 4, 2 and 1, respectively. All results are obtained specifying the parameters as $D = 8$, $q^* = 5$ and $t^* = 3$. We summarize the results of the simulation in terms of bias and variance of the estimate. For a given point x , due to the symmetric situation, both statistics mainly depend on the distance $d(x, x_c)$ of x to the center x_c of the image.

Figure 7 illustrates the dependence of bias and standard deviation of the estimates from $d(x, x_c)$ for image 1 (upper row) and image 2 (lower row) for the different values of σ . We restrict the presentation to the interesting region near the boundary. Both bias and standard deviation are negligible outside these regions. In general, for a gridpoint $x^0 \in A_m$, bias depends on the probability $\mathbf{P}(\hat{U}(x^0) \not\subseteq A_m)$, being negligible if this quantity is small. Variance depends on the mean size of $\hat{U}(x^0)$ and again on $\mathbf{P}(\hat{U}(x^0) \not\subseteq A_m)$. In case of $\sigma = .125$ for most x^0 all $U(x^0) \not\subseteq A_m$ are rejected, resulting in a vanishing bias for interior points even near the boundary.

Note that for both images and $\sigma \leq .5$ bias and variance of the estimate are very small for all points x with $d(x, x_c) > 12\delta$ and $d(x, x_c) > 24\delta$, respectively. Additionally, for $\sigma = 1$ bias and variance can be reduced in the outer regions by simply increasing D , see also Figure 6.

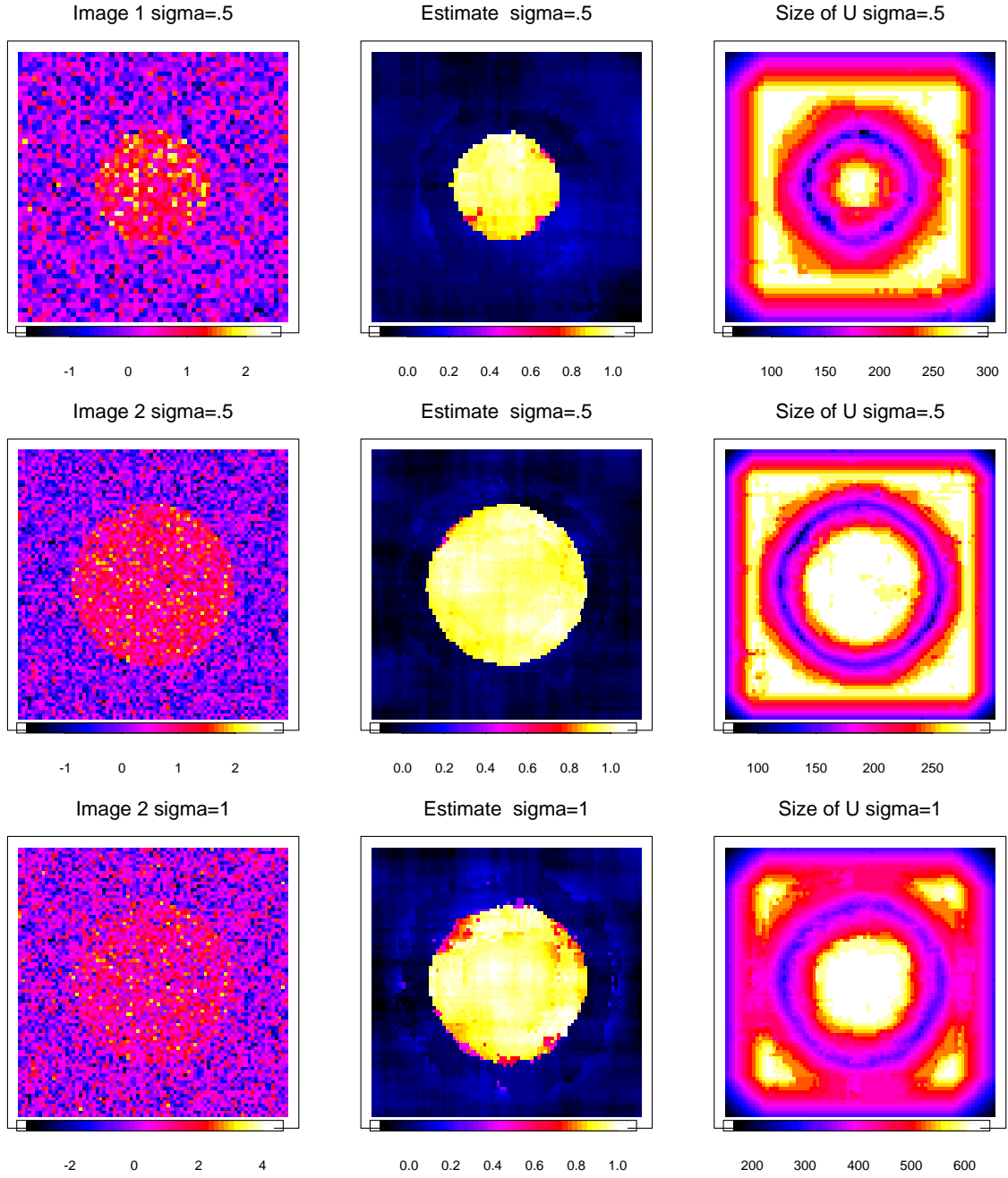


FIGURE 6. Distorted images, estimates and size of selected sets U for three situations. First row: circle of radius 12δ , signal/noise ratio 2, estimate with $D = 8$ and $q^* = 5$. Second row: circle of radius 24δ , signal/noise ratio 2, estimate with $D = 8$ and $q^* = 5$. Third row: circle of radius 24δ , signal/noise ratio 1, estimate with $D = 12$ and $q^* = 5$.

For $d(x, x_c) < 12\delta$ and $d(x, x_c) < 24\delta$ we observe the effects of orientation and curvature of the boundary g . Curvature of g mainly restricts the size of possible sets $U(x^0)$ inside the circle, therefore leading to poorer results near the boundary g and for image 1.

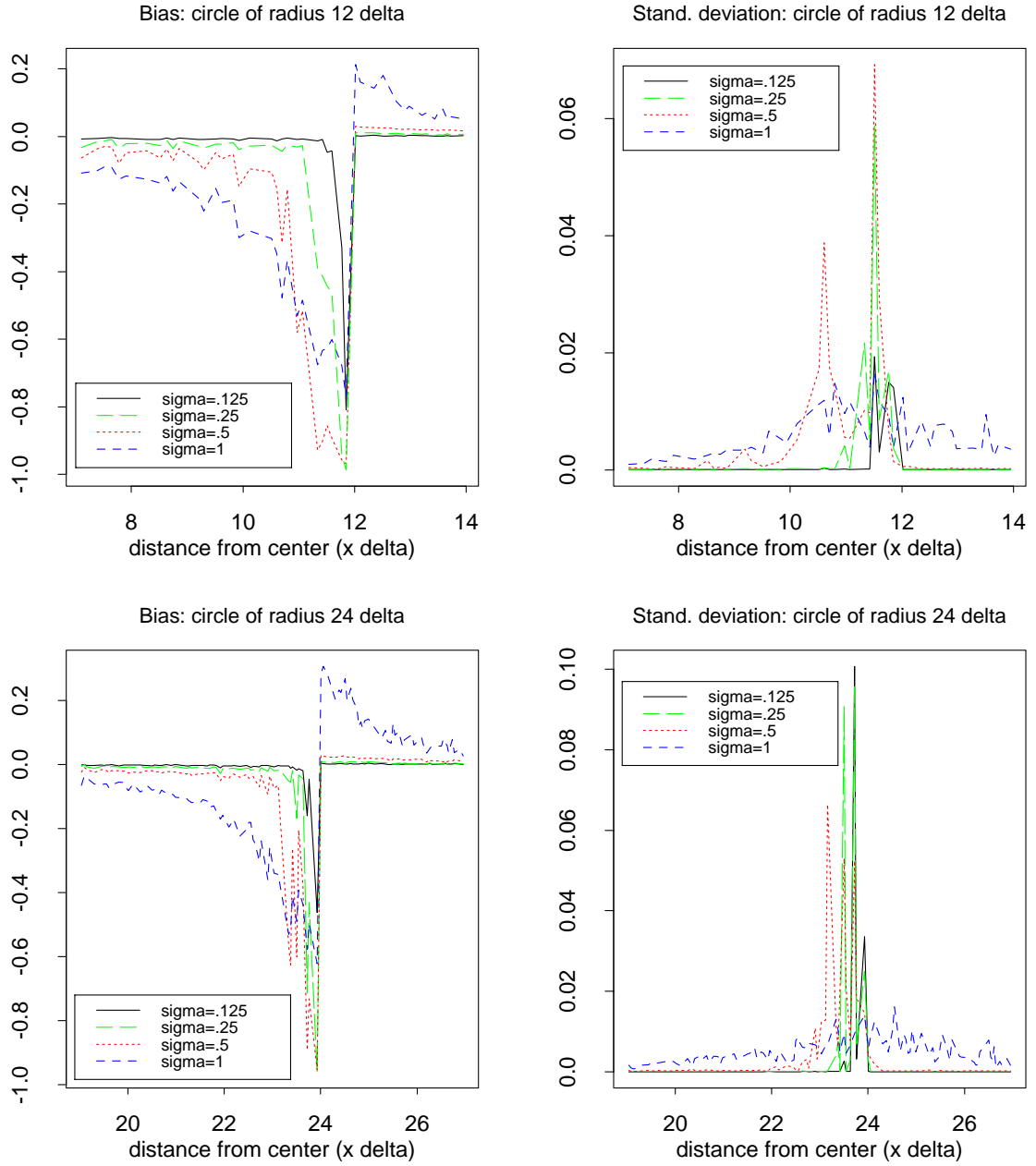


FIGURE 7. Estimated bias and variance for image 1 (circle of radius 12δ , upper row) and image 2 (circle of radius 24δ , lower row) as functions of the distance from the center. Signal to noise ratio equal to 8, 4, 2 and 1.

We also observe the effect of expansion from the side of a larger region w.r.t. the smaller one: this means that the interior points near the boundary can be estimated as 0. This leads to the negative bias inside. The influence of orientation is reflected in the roughness of the bias curves, showing smaller values for preferable orientations.

Proofs

In this section we present the proofs of Theorem 3.1 through 4.4.

5.1. Proof of Theorem 3.1

We begin with some preliminary results. An “ideal” window for estimating $f(x^0)$ coincides clearly with the region A_m containing x^0 . Hence the idea of the proposed procedure is to select adaptively the largest window among the considered class \mathcal{U} which is contained in A_m . A necessary property of every such procedure is to accept each window contained in A_m with a high probability. Our first result shows that the previously described procedure possesses this properties.

Proposition 5.1. *Let $x^0 \in A_m$ for some $m = 1, \dots, M$ and let $U \in \mathcal{U}$ be such that $U \subseteq A_m$. Then*

$$\mathbf{P}_f(\varrho_U = 1) \leq e^{-\lambda}.$$

Proof. Let some U with the property $U \subseteq A_m$ be fixed and let $V \in \mathcal{V}(U)$. The function f is constant on U and hence on V . Using the model equation (1.1) we obtain

$$T_{U,V} = \sigma_{U,V}^{-1} \left(\frac{1}{N_V} \sum_V \xi_i - \frac{1}{N_U} \sum_U \xi_i \right).$$

Obviously we have $\mathbf{E}_f T_{U,V} = 0$. Recall now that the factor $\sigma_{U,V}$ was defined as the standard deviation of the stochastic term of the difference $\hat{f}_V - \hat{f}_U$. Hence $\mathbf{E}_f T_{U,V}^2 = 1$. Since $T_{U,V}$ is a linear combination of Gaussian variables ξ_i , $T_{U,V}$ itself is Gaussian with zero mean and the unit variance i.e. standard normal. Now

$$\mathbf{P}_f(|T_{U,V}| > t^*) \leq \exp\{-\lambda - \log C_U\} = e^{-\lambda} C_U^{-1}.$$

This estimate and condition (2.1) allow to bound the probability of rejecting U in the following way

$$\mathbf{P}_f(\varrho_U = 1) \leq \sum_{V \in \mathcal{V}(U)} \mathbf{P}_f(|T_{U,V}| > t^*) \leq C_U e^{-\lambda} C_U^{-1} = e^{-\lambda}.$$

□

The next statement can be viewed as a complement to Proposition 5.1. We consider now the case of a “bad” window containing two non-intersecting subwindows V_1 and V_2 with different values of the image function f . The result says that such a window will be rejected with a high probability.

Proposition 5.2. *Let $U \in \mathcal{U}$ and let $V_1, V_2 \in \mathcal{V}(U)$ be such that the function f is constant within each V_j ,*

$$f(x) = a_j, \quad x \in V_j, \quad j = 1, 2.$$

Denote

$$s_{V_1, V_2} = \sigma \sqrt{N_{V_1}^{-1} + N_{V_2}^{-1}}.$$

If

$$|a_1 - a_2| \geq (\sigma_{U,V_1} + \sigma_{U,V_2} + s_{V_1,V_2}) t^* \quad (5.1)$$

with $\sigma_{U,V}^2 = \sigma(N_V^{-1} - N_U^{-1})$, then

$$\mathbf{P}_f(\varrho_U = 0) \leq e^{-\lambda}/C^*.$$

Remark 5.1. In view of the trivial inequalities $\sigma_{U,V} \leq \sigma N_V^{-1/2}$ and $\sqrt{N_{V_1}^{-1} + N_{V_2}^{-1}} \leq N_{V_1}^{-1/2} + N_{V_2}^{-1/2}$, condition (5.1) is fulfilled if

$$|a_1 - a_2| \geq 2\sigma t^* (N_{V_1}^{-1/2} + N_{V_2}^{-1/2}). \quad (5.2)$$

Proof. By definition

$$\mathbf{P}_f(\varrho_U = 0) \leq \mathbf{P}_f(\varrho_{U,V_1} = \varrho_{U,V_2} = 0).$$

Next, the event $\{\varrho_{U,V} = 0\}$ means $|T_{U,V}| \leq t^*$ or equivalently

$$|\hat{f}_U - \hat{f}_V| \leq \sigma_{U,V} t^*.$$

This yields

$$|\hat{f}_{V_1} - \hat{f}_{V_2}| \leq (\sigma_{U,V_1} + \sigma_{U,V_2}) t^*.$$

Now using the fact that $V_1 \cap V_2 = \emptyset$, we get the following decomposition, cf. the proof of Proposition 5.1,

$$\hat{f}_{V_1} - \hat{f}_{V_2} = a_1 - a_2 + N_{V_1}^{-1} \sum_{V_1} \xi_i - N_{V_2}^{-1} \sum_{V_2} \xi_i = a_1 - a_2 + s_{1,2} \zeta_{1,2}$$

where $\zeta_{1,2}$ is a standard normal random variable. Therefore,

$$\begin{aligned} \mathbf{P}_f(\varrho_U = 0) &\leq \mathbf{P}(|a_1 - a_2 + s_{1,2} \zeta_{1,2}| \leq (\sigma_{U,V_1} + \sigma_{U,V_2}) t^*) \\ &\leq \mathbf{P}(s_{1,2} |\zeta_{1,2}| \geq |a_1 - a_2| - (\sigma_{U,V_1} + \sigma_{U,V_2}) t^*). \end{aligned}$$

Using the condition of the proposition, we obtain

$$\mathbf{P}_f(\varrho_U = 0) \leq \mathbf{P}(|\zeta_{1,2}| > t^*) \leq e^{-\lambda - \log C^*} = e^{-\lambda}/C^*$$

and the assertion follows. \square

We need one more result concerning the situation when a window U from \mathcal{U} is not entirely contained in A_m but there is its subwindow V which is in A_m .

Proposition 5.3. *Let $x^0 \in A_m$, $U \in \mathcal{U}$ and let V from $\mathcal{V}(U)$ be such that $V \subseteq A_m$. If $\varrho_{U,V} = 0$, then the difference $|\hat{f}_U - f(x^0)|$ can be estimated in the following way: for any $z \geq 1$*

$$\mathbf{P}(|\hat{f}_U - f(x^0)| > \sigma N_V^{-1/2}(z + t^*), \varrho_{U,V} = 0) \leq \exp\{-z^2/2\}.$$

Proof. The event $\{\varrho_{U,V} = 0\}$ means that $|\hat{f}_U - \hat{f}_V| \leq \sigma_{U,V} t^*$. Therefore,

$$|\hat{f}_U - a_m| \leq |\hat{f}_U - \hat{f}_V| + |\hat{f}_V - a_m| \leq \sigma_{U,V} t^* + |\hat{f}_V - a_m|.$$

Next, $\sigma_{U,V} \leq \sigma N_V^{-1/2}$ and $\zeta = \sigma^{-1} N_V^{1/2}(\hat{f}_V - a_m)$ is a standard Gaussian random variable, see the proof of Proposition 5.1. This gives

$$\mathbf{P}_f(|\hat{f}_U - f(x^0)| > \sigma N_V^{-1/2}(z + t^*), \varrho_{U,V} = 0) \leq \mathbf{P}(|\zeta| > z) \leq \exp\{-z^2/2\}$$

as required. \square

Now we turn directly to the proof of Theorem 3.1. First of all, since U^* is contained in A_m , due to Proposition 5.1 the window U^* will be rejected only with a very small probability, namely

$$\mathbf{P}_f(\varrho_{U^*} = 1) \leq e^{-\lambda}.$$

Since obviously for every $z > 0$ by Proposition 5.1

$$\mathbf{P}_f\left(|\hat{f}(x^0) - f(x^0)| > z, \varrho_{U^*} = 1\right) \leq \mathbf{P}_f(\varrho_{U^*} = 1) \leq e^{-\lambda}$$

it suffices to consider only the situation when U^* is accepted i.e. $\varrho_{U^*} = 0$.

Let window \hat{U} be selected by the procedure. Then $\varrho_{\hat{U}} = 0$ and, under the case of $\varrho_{U^*} = 0$, by definition of \hat{U} ,

$$N_{\hat{U}} \geq N_{U^*}.$$

Next, due to condition (U.2), there is a subwindow V in $\hat{U} \cap U^*$ with at least K design points which is therefore contained in A_m . If \hat{U} contains also another subwindow V' with $N_{V'} \geq K$ which lies outside A_m , then we observe by Proposition 5.2, see also Remark 5.1, that the probability to accept \hat{U} is very small. Namely, let \mathcal{U}' be the subset in \mathcal{U} of all U with this property. Then

$$\mathbf{P}_f(\varrho_{\hat{U}} = 0) \leq \sum_{U \in \mathcal{U}'} \mathbf{P}_f(\varrho_U = 0) \leq \sum_{U \in \mathcal{U}'} e^{-\lambda}/C^* \leq e^{-\lambda},$$

and arguing as above we reduce our consideration to the case when $\hat{U} \setminus A_m$ does not contain any such V' . Let \mathcal{U}'' be the subset in \mathcal{U} of all windows U with the last property. By condition (U.3), for each $U \in \mathcal{U}''$, there is $V \in \mathcal{V}(U)$ such that $V \subseteq \hat{U} \cap A_m$ and $N_V \geq 0.5N_{U^*}$. We denote this V by $V(U)$. The definition of \hat{U} ensures that $\varrho_{\hat{U},V} = 0$ and we conclude using Proposition 5.3 with $z = t^*$

$$\begin{aligned} \mathbf{P}_f\left(|\hat{f}_{\hat{U}} - f(x^0)| > 2\sigma(N_{U^*}/2)^{-1/2}t^*\right) \\ \leq 2e^{-\lambda} + \sum_{U \in \mathcal{U}''} \mathbf{P}_f\left(|\hat{f}_U - f(x^0)| > 2\sigma N_{V(U)}^{-1/2}t^*, \varrho_{U,V(U)} = 0\right) \\ \leq 2e^{-\lambda} + \sum_{U \in \mathcal{U}''} e^{-\lambda - \log C^*} \\ \leq 2e^{-\lambda} + C^* e^{-\lambda - \log C^*} = 3e^{-\lambda} \end{aligned}$$

and the assertion follows.

5.2. Proof of Theorem 4.2

The statement of this theorem is a direct application of Theorem 3.1. The main problem is to verify that there is a window U in $\mathcal{U}_{\mathcal{D},K}$ with at least d_γ^2 points which lies in A . Then automatically $N_{U^*} \geq d_\gamma^2$.

Let $z = g'(x_1^0)$. We known that $|z| \leq 1$. To be more definitive, we suppose that $0 \leq z \leq 1$. The case of a negative z can be considered in the same way. We

denote also

$$\Delta = (n/K)^{-\gamma/(\gamma+1)} P^{1/(\gamma+1)}$$

and $y = x_2^0 + \Delta$.

Lemma 5.1. *For all x_1 with $|x_1 - x_1^0| \leq \delta d_\gamma$*

$$y + z(x_1 - x_1^0) \leq g(x_1).$$

Proof. The smoothness condition $g \in \Sigma(\gamma, P)$ implies for all $h > 0$

$$\sup_{|t| \leq h} |g(x_1^0 + t) - g(x_1^0) - zt| \leq Ph^\gamma.$$

Therefore

$$g(x_1^0 + t) \geq g(x_1^0) + zt - Ph^\gamma = y + zt + \Delta - Ph^\gamma$$

for all $|t| \leq h$. Now we apply $h = \delta d_\gamma$ and the assertion follows because of

$$Ph^\gamma = P(\delta d_\gamma)^\gamma \leq P^{1/(\gamma+1)} (\delta^2 K)^{\frac{\gamma}{\gamma+1}} = \Delta.$$

□

Now we define the required window U using the line L passing through (x_1^0, y) with the angle z . For this line we have the equation

$$x_2 - y = z(x_1 - x_1^0).$$

By Lemma 5.1 this line lies under the curve G at least within the square Q_{d_γ} . Now we define U as the window from \mathcal{U}_{d_γ} corresponding to the line L , $U = U_{d_\gamma, L}$.

Obviously this window is a subset of A_m and it remains only to check that $K_{d, L} \geq K$. Then we have $U \in \mathcal{U}_{d_\gamma, K} \subset \mathcal{U}$.

We use the following technical

Lemma 5.2. *Let z be any number with $0 \leq z \leq 1$. Then for every positive number v there is a rational number p/q with $0 \leq p \leq q \leq v$ such that*

$$|zq - p| \leq v^{-1}.$$

Proof. Suppose without loss of generality that z is an irrational number from the interval $[0, 1]$. Denote by $(p_k/q_k)_{k \geq 1}$ the sequence of rational numbers which gives the best rational approximation of z , see Khintchine (1949). It can be defined as a sequence of continued fractions: we begin with $r_0 = z^{-1}$ and define inductively $n_k = \lfloor r_{k-1} \rfloor$, $r_k = (r_{k-1} - n_k)^{-1}$ for $k = 1, 2, \dots$; then p_k/q_k can be described as the following continued fraction

$$\frac{p_k}{q_k} = \frac{1}{n_1 + \frac{1}{n_2 + \cdots \frac{1}{n_{k-1} + \frac{1}{n_k}}}}.$$

This approximation has the following properties, Khintchine (1949, Section 3,4):

$$\left| z - \frac{p_k}{q_k} \right| \leq \frac{1}{q_k q_{k+1}}, \quad (5.3)$$

Given a number v , denote

$$k^* = \max\{k : q_k \leq v\}$$

so that $q_{k^*+1} > v$. By (5.3), $|zq_{k^*} - p_{k^*}| \leq 1/q_{k^*+1} < 1/v$ and the assertion follows. \square

We now apply this lemma with $z = g'(x_1^0)$ taking first $v_K = d_\gamma/K$. Let p_K, q_K be the corresponding numbers with $p \leq q \leq d_\gamma/K$ such that

$$|q_K z - p_K| \leq K/d_\gamma.$$

If it holds also

$$|q_K z - p_K| \leq (K-1)/d_\gamma$$

then we set $q_{K-1} = q_K$ and $p_{K-1} = p_K$. Otherwise there are other q_{K-1} and p_{K-1} with the properties $p_{K-1} \leq q_{K-1} \leq d_\gamma/(K-1)$ and

$$|q_{K-1} z - p_{K-1}| \leq (K-1)/d_\gamma.$$

We continue in this way taking each time either $(p_{k-1}, q_{k-1}) = (p_k, q_k)$ or defining a new pair (p_{k-1}, q_{k-1}) with the properties $p_{k-1} \leq q_{k-1} \leq d_\gamma/(k-1)$ and

$$|q_{k-1} z - p_{k-1}| \leq (k-1)/d_\gamma,$$

$k = K-1, K-2, \dots, 1$. By construction all pairs (p_k, q_k) verify $p_k \leq q_k \leq d_\gamma$ and

$$|q_k z - p_k| \leq K/d_\gamma.$$

Denote now for every $k = 1, \dots, K$

$$X^{(k)} = \begin{cases} (x_1^0 + q_k \delta, x_2^0 + p_k \delta), & \text{if } q_k z \leq p_k, \\ (x_1^0 - q_k \delta, x_2^0 - p_k \delta) & \text{otherwise.} \end{cases}$$

Obviously all $X^{(k)}$ are grid points. Moreover, since $p_k \leq q_k \leq d_\gamma$, we have $X^{(k)} \in Q_{d_\gamma}$. We aim to show that all $X^{(k)}$ lie between L and L' where L' is the line with the equation $x_2 - x_2^0 = z(x_1 - x_1^0)$. Suppose for simplicity that $q_k z \leq p_k$ and check $X^{(k)} \in U_{d_\gamma, L}$. Indeed

$$x_2^0 + p_k \delta - y - z(x_1^0 + q_k \delta - x_1^0) = \Delta + (p_k - q_k z) \delta \leq \Delta - \delta K/d_\gamma \leq 0$$

which means exactly that $X^{(k)}$ is under the line L and therefore, $X^{(k)} \in U_L$. Further

$$x_2^0 + p_k \delta - x_2^0 - z(x_1^0 + q_k \delta - x_1^0) = \delta(p_k - zq_k) \geq 0$$

and hence $X^{(k)}$ is over L' as required. The case of $q_k z > p_k$ can be considered in the same way.

We have checked that $U = U_{d_\gamma, L}$ is in \mathcal{U} and at the same time $U \subset A_m$. This means that the set of windows from \mathcal{U} with these properties is not empty and the “ideal” window U^* satisfies $N_{U^*} \geq N_U \geq d_\gamma^2$.

Since the set A_m has the structure of a boundary fragment within the square $|x - x^0| \leq \epsilon$, and since all the considered windows from \mathcal{U} lie inside this square (because $D \leq \epsilon n^{1/2}$), condition (U.3) is clearly fulfilled. Now we may apply Theorem 3.1 which leads exactly to the desirable statement.

5.3. Proof of Theorem 4.3

The proof of this results can be derived along the same line as the proof of Theorem 4.2 and is even simpler. Indeed, we may take the line L passing through x^0 with the angle $z = p/q$. Then this line passes also through the design points $X^{(k)} = (x_1^0 + kq, x_2^0 + kp)$ for all integer k . Then the interval between the points $X^{(-K)}$ and $X^{(K)}$ on this line contains at least $2K + 1$ design points and therefore the window $U_{d,L}$ with $d = Kq$ is in \mathcal{U} . The condition (4.4) provides that this window is also in A_m and we end up similarly to Theorem 4.2.

5.4. Proof of Theorem 4.4

Different methods for obtaining the lower bound results in edge estimation are presented in Korostelev and Tsybakov (1993). We cannot apply these methods directly since they are developed for a random design and we operate with the regular design. But we follow the same route and we therefore present only a sketch of the proof concentrating on the points specific for our situation.

Let some γ from the interval $(1, 2]$ and some integers K, D be fixed. Let also $z = p/q$ be a unreducible rational number with $p \leq q \leq D$. Set

$$h = \min\{qK\delta, (\delta/q)^{1/\gamma}\},$$

where $\delta = n^{-1/2}$.

Let now ϕ be a smooth function satisfying the conditions

- (a) ϕ is symmetric and nonnegative;
- (b) $\phi(0) = \sup_t \phi(t)$ and $0 < \phi(0) \leq 1$;
- (c) ϕ is compactly supported on $[-1, 1]$;
- (d) ϕ belongs to the Hölder ball $\Sigma(\gamma, 1)$.

Denote

$$\phi_h(t) = h^\gamma \phi(t/h).$$

Then (d) ensures that $\phi_h \in \Sigma(\gamma, 1)$ for all $h > 0$. Next, set

$$\begin{aligned} g_1(x_1) &= (x_1 - x_1^0)p/q - \phi_h(0)/2 \\ g_2(x_2) &= (x_1 - x_1^0)p/q + \phi_h(0)/2 - \phi_h(x_1 - x_1^0). \end{aligned}$$

Each function g_k determines the boundary fragment A_k with the edge G_k ,

$$\begin{aligned} A_k &= \{x = (x_1, x_2) : x_2 \leq g_k(x_1)\}, \\ G_k &= \{x = (x_1, x_2) : x_2 = g_k(x_1)\}, \quad k = 1, 2. \end{aligned}$$

Set also

$$B = A_2 \setminus A_1 = \{x = (x_1, x_2) : g_1(x_1) < x_2 \leq g_2(x_1)\}.$$

Below we make use of the following technical assertion.

Lemma 5.3. *The following assertions hold*

- (i) $g_1, g_2 \in \Sigma(\gamma, 1)$ and $g_1'(x_1^0) = g_2'(x_1^0) = q/p$;
- (ii) $|g(x_1^0) - x_2^0| \geq \kappa h^\gamma$ for some $\kappa > 0$ depending on ϕ only;

(iii) The number N of design points in the set B is at most $2K - 1$,

$$N = \#\{X_i \in B\} \leq 2K - 1.$$

Proof. Assertions (i) and (ii) are obvious. We comment on (iii).

Let L be the line passing through x^0 with the angle z , i.e. L is described by the equation $x_2 - x_2^0 = z(x_1 - x_1^0)$. We fix also two points $x^- = (x_1^0 - Kq\delta, x_2^0 - Kp\delta)$ and $x^+ = (x_1^0 + Kq\delta, x_2^0 + Kp\delta)$ on this line. Since $h \leq qK\delta$, then the interval passes exactly through $2K - 1$ design points. We intend to show that there is no other design points in B that implies the assertion in view of property (c) of ϕ .

Let $x = (x_1, x_2)$ be a design point with coordinates $(x_1^0 + q'\delta, x_2^0 + p'\delta)$ such that $p'/q' \neq p/q$. To verify that $x \notin B$, it suffices to check that

$$|p'\delta - q'\delta p/q| > |\phi_h(q'\delta) - \phi_h(0)/2|.$$

Since $p'/q' \neq p/q$, then

$$|p' - q'p/q| = q^{-1}|p'q - q'p| \geq q^{-1}$$

and hence $|p'\delta - q'\delta p/q| \geq \delta/q$. In view of (b), we have $\phi_h(x_1 - x_1^0) \leq \phi_h(0) \leq h^\gamma$ and by definition of h we have $h^\gamma \leq \delta/q$ and (iii) follows. \square

Denote $f_k(x) = \mathbf{1}(x \notin A_k) = \mathbf{1}(x_2 > g_k(x_1))$ for $x = (x_1, x_2)$, $k = 1, 2$. Note that $f_1(x^0) = 0$ and $f_2(x^0) = 1$. Now for any estimator $\tilde{f}(x^0)$

$$\begin{aligned} R &:= \mathbf{P}_1 \left(|\tilde{f}(x^0)| > 1/2 \right) + \mathbf{P}_2 \left(|\tilde{f}(x^0) - 1| > 1/2 \right) \\ &= \mathbf{E}_1 \left\{ \mathbf{1} \left(|\tilde{f}(x^0)| > 1/2 \right) + Z \mathbf{1} \left(|\tilde{f}(x^0) - 1| > 1/2 \right) \right\} \end{aligned} \quad (5.4)$$

where \mathbf{E}_k stands for E_{g_k} , $k = 1, 2$, and $Z = dP_2/dP_1$. It is easy to show that the optimal decision $\tilde{f}(x^0)$ for the latter two-point problem is of the form $\tilde{f}(x^0) = \mathbf{1}(Z \geq 1)$ and hence

$$R \geq \mathbf{E}_1 \mathbf{1}(Z \geq 1) = \mathbf{P}_1(Z \geq 1).$$

Next, making use of the model equation (1.1) we get the following representation of the likelihood Z ,

$$Z = \exp \left\{ \sigma^{-2} \sum_B \xi_i - \frac{N\sigma^{-2}}{2} \right\}$$

where the sum over B means the sum over design points X_i falling in B and the random errors ξ_i are normal $\mathcal{N}(0, \sigma^2)$. If we set

$$\zeta = \frac{1}{\sigma\sqrt{N}} \sum_B \xi_i,$$

then Lemma 5.3, (ii) and (iii) implies that ζ is under P_1 a standard normal random variable and

$$\begin{aligned} P_1(Z > 1) &= P_1\left(\exp\left\{\sigma^{-1}\sqrt{N}\zeta - \sigma^{-2}N/2\right\} > 1\right) \\ &= P_1\left(\zeta > \sigma^{-1}\sqrt{N}/2\right) \\ &\leq P_1\left(\zeta > \sigma^{-1}\sqrt{K/2}\right) \\ &= 1 - \Phi\left(\sigma^{-1}\sqrt{K/2}\right) > 0 \end{aligned}$$

where Φ is the Laplace function and the required assertion follows.

References

- [1] Blake, A. and Zisserman, A. (1987). *Visual reconstruction*. MIT Press, Cambridge, MA.
- [2] Donoho, D.L. (1996). ??
- [3] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images. *IEEE Trans. on PAMI* **6** 721–741.
- [4] Girard, D. (1990). From template matching to optimal approximation by piecewise smooth curves. In: *Curves and Surfaces in Computer Vision and Graphics*, Proc. Conf. Santa Clara, Calif., 1990, 174–182.
- [5] Grenander, U. (1976, 1978, 1981). *Lectures in Pattern Theory*, vol. 1–3. Springer, New York. (Applied Math. Sciences, vol. 18, 24, 33).
- [6] Haralick, R.M. (1980). Edge and region analysis for digital image data. *Comput. Graphics and Image Processing* **12** 60–73.
- [7] Huang, J.S. and Tseng, D.H. (1988). Statistical Theory of edge detection. *Graphics and Image Processing* **43** 337–346.
- [8] Khintchine (1949) *Continued fractions*. ITTL, Moscow.
- [9] Korostelev, A. and Tsybakov, A. (1993). *Minimax Theory of Image Reconstruction*. Springer Verlag, New York–Heidelberg–Berlin.
- [10] Lee, J.S. (1983). Digital image smoothing and the sigma-filter. *Computer Vision, Graphics and Image Processing* **24** 255–269.
- [11] Lepskii, O.V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35**, no. 3, 459–470.
- [12] Lepskii, O.V. (1992). Asymptotic minimax adaptive estimation. 2. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* **37**, no. 3, 468–481.
- [13] Lepski, O., Mammen, E. and Spokoiny, V. (1997). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection. *Annals of Statistics*, **25**, no. 3, 929–947.
- [14] Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, **25**, no. 6, 2512–2546.
- [15] Mammen, E. and Tsybakov, A. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23**, no. 2, 502–524.
- [16] Marr, D. (1982). *Vision*. Freeman and Co., San Francisco.
- [17] Müller, H.G. and Song, K.S. (1994). Maximum estimation of multidimensional boundaries. *J. Multivariate Anal.* **50**, no.2, 265–281.
- [18] Mumford, D. and Shah, J. (1989). Optimal approximation by piecewise smooth functions and associated variational problem. *Comm. Pure Appl. Math.* **42** 577–685.
- [19] Petrov, V.V. (1975) *Sums of Independent Random Variables*. Springer, New York.
- [20] Pratt, W.K. (1978). *Digital Image Processing*. J.Wiley, New York.
- [21] Ripley, B.D. (1988). *Statistical Inference for Spatial Processes*. Cambridge Univ. Press.
- [22] Rosenfeld, A. and Kak, A.C. (1982). *Digital Picture Processing*. Academic Press, London.

- [23] Shiau, J.J. Wahba, G. and Johnson, D.R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise two- or three-dimensional objective analysis. *J. Atmospheric and Oceanic Technology* **3** 714–725.
- [24] Spokoiny, V. (1996). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. Preprint **291**, Weierstrass Institute, Berlin. *Ann. Statist.*, tentatively accepted.
- [25] Titterton, D.M. (1985). Common structure of smoothing technique in statistics. *Intern. Stat. Review* **53** 141–170.
- [26] Tsybakov, A. (1989). Optimal estimation accuracy of nonsmooth images. *Problem Inf. Trans.*, **25**, 180–191.

WEIERSTRASS INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTR. 39,
10117 BERLIN, GERMANY

E-mail address: `polzehl@wias-berlin.de`

WEIERSTRASS INSTITUTE FOR APPLIED ANALYSIS AND STOCHASTICS, MOHRENSTR. 39,
10117 BERLIN, GERMANY

E-mail address: `spokoiny@wias-berlin.de`